CISC 371 Class 6

Stationarity and the Hessian Matrix

Texts: [1] pp. 23–34; [2] pp. 27–51

Main Concepts:

- *Stationarity is $D_{\vec{v}} f(\vec{w}^*) = 0$ for all $\vec{v}$*
- *Stationarity is necessary for minimization*
- *Hessian matrix $\underline{\nabla}^2 f$ or H_f*
- *Eigenvalues of the Hessian matrix*
- *Saddle points*

Sample Problem, Data Analytics: Where, in a temperature map, is the temperature "steady"?

In Class 1, we defined the meaning of a vector $\vec{w}^*$ being a minimizer of a function $f(\vec{w})$. In Class 4, we defined the meaning of a scalar argument t^* being a stationary point of a function $f(t)$. We will now explore stationarity for a function that has a vector argument, and the extension of the concept of a second derivative for such a function.

6.1 Stationary Point of a Function with a Vector Argument

In plain English, a *stationary point* of $f(\vec{w})$ is a point $\vec{w}^*$ where the function f is "flat". Of the multiple equivalent definitions, we will mean that, at a stationary point, the directional derivative is zero in every direction. There is a technical requirement – shared by all definitions – that the stationary point must be an interior point; for now, we will use the definition in Section 3.1 for functions that are defined everywhere in a vector space $\mathbb{R}^n$:

$\vec{w}^*$ is a *stationary point* of $f(\vec{w})$ is defined as: for all $\vec{v} \in \mathbb{R}^n$, $D_{\vec{v}} f(\vec{w}^*) = 0$

For this definition, a simple theorem is that the gradient is zero at a stationary point. This version of stationarity is, in 1-form notation,

$\vec{w}^*$ is a *stationary point* of $f(\vec{w})$ if and only if $\underline{\nabla} f(\vec{w}^*) = \underline{0}$

The vector statement, which is commonly used as the definition of stationarity, is

$\vec{w}^*$ is a *stationary point* of $f(\vec{w})$ if and only if $[\underline{\nabla} f(\vec{w}^*)]^T = \vec{0}$

The entry-wise statement, which is also commonly used as the definition of stationarity, is

$$\vec{w}^* \text{ is a } \textit{stationary point} \text{ of } f(\vec{w}) \text{ if and only if, for all } 1 \le j \le n,\ \frac{\partial f(\vec{w}^*)}{\partial w_j} = 0$$

Our definition and theorems on stationarity all share an implication. At a stationary point $\vec{w}^*$, an infinitesimal change of $\vec{w}^*$ in any direction produces zero change in the value of f.

In a vector space, as for a scalar argument, a necessary condition for a point to be a strict minimizer – or to be a minimizer – is that it is a a stationary point. Also as for a scalar argument, this property is not a sufficient condition. To explore stationarity further, we need to extend the concept of second derivative from a scalar argument to a vector space.

6.2 The Hessian Matrix of a Function

For a function with a scalar argument, which is written as $f : \mathbb{R} \to \mathbb{R}$ or as $f(t)$, the second derivative is variously written as $f''(t)$ or as $\frac{d^2 f}{dt^2}(t)$. For a function with a vector argument, which is written as $f : \mathbb{R}^n \to \mathbb{R}$ or as $f(\vec{w})$, there is no single second derivative. We can easily write the combined second derivatives in two ways: as individual entries or as a Jacobian matrix.

As entries, the function $f(\vec{w})$ can be differentiated first by entry j of the vector $\vec{w}$ and then by entry i if $\vec{w}$. Abbreviating the result of the differentiations as h_{ij}, we can write

$$h_{ij} \stackrel{\text{def}}{=} \frac{\partial^2}{\partial w_i \partial w_j} \tag{6.1}$$

The values h_{ij} can be interpreted as being entries of a $n \times n$ matrix, which we will write as $\underline{\nabla}^2 f$. Because a fundamental result of basic calculus is that order of differentiation does not matter, we can infer that $h_{ij} = h_{ji}$ so the matrix must be symmetric. The matrix with these entries is defined as the *Hessian matrix*, which is

$$\underline{\nabla}^2 f(\vec{w}_0) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial^2 f}{\partial^2 w_1}(\vec{w}_0) & \frac{\partial^2 f}{\partial w_1 \partial w_2}(\vec{w}_0) & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_m}(\vec{w}_0) \\ \frac{\partial^2 f}{\partial w_2 \partial w_1}(\vec{w}_0) & \frac{\partial^2 f}{\partial^2 w_2}(\vec{w}_0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_n \partial w_1}(\vec{w}_0) & \frac{\partial^2 f}{\partial w_n \partial w_2}(\vec{w}_0) & \cdots & \frac{\partial^2 f}{\partial^2 w_n}(\vec{w}_0) \end{bmatrix} \tag{6.2}$$

Because the second partial derivative is a symmetric operator, the Hessian matrix is symmetric.

Another way to write the Hessian matrix is to use the derivative of the transpose of the gradient of f. The gradient of f is a 1-form, so the transpose of the gradient is a vector. The j^{th} entry of the vector is the function $\dfrac{\partial}{\partial w_j} f(\vec{w})$. We can find the Jacobian matrix of the transpose, the entries of which correspond to the entries of the Hessian matrix defined in Equation 6.2. This definition is

$$\underline{\nabla}^2 f(\vec{w}_0) \stackrel{\text{def}}{=} \underline{\nabla}\,[\underline{\nabla}\, f(\vec{w}_0)]^T = \underline{\nabla} \begin{bmatrix} \dfrac{\partial f}{\partial w_1}(\vec{w}_0) \\ \dfrac{\partial f}{\partial w_2}(\vec{w}_0) \\ \vdots \\ \dfrac{\partial f}{\partial w_n}(\vec{w}_0) \end{bmatrix} \tag{6.3}$$

The equivalence of Equation 6.3 and Equation 6.2 arises in part from the symmetry of partial differentiation.

The Hessian matrix is symmetric, and each entry is a real number, so the eigenvalues are real number and the eigenvectors are in the vector space $\mathbb{R}^n$. We know, from elementary linear algebra, that there are five characterizations of symmetric matrices that can be distinguished by using the eigenvalues. For a symmetric matrix $H = H^T$, these are:

- $\underline{\nabla}^2 f$ is *positive definite*, written $]\underline{\nabla}^2 f] \succ 0$, where each eigenvalue $\lambda_i > 0$
- $\underline{\nabla}^2 f$ is *positive semidefinite*, written $[\underline{\nabla}^2 f] \succeq 0$, where each eigenvalue $\lambda_i \geq 0$
- $\underline{\nabla}^2 f$ is *indefinite* if there is some eigenvalue $\lambda_i > 0$ and some eigenvalue $\lambda_j < 0$
- $\underline{\nabla}^2 f$ is *negative definite*, written $[\underline{\nabla}^2 f] \prec 0$, where each eigenvalue $\lambda_i < 0$
- $\underline{\nabla}^2 f$ is *positive semidefinite*, written $[\underline{\nabla}^2 f] \preceq 0$, where each eigenvalue $\lambda_i \leq 0$

In Class 4, we observed that there were three kinds of stationary points for a function with a scalar argument. These were associated with the sign of the second derivative at the stationary point t^*. How are eigenvalues of the Hessian matrix associated with the possible kinds of stationary points?

There are three kinds of eigenvalue combinations for which we can make a sure decision about the objective function, and two kinds for which we must remain uncertain. The combinations of eigenvalues are:

$\boldsymbol{H} \succ 0$: $\vec{w}^*$ is a strict local minimizer

$\boldsymbol{H} \prec 0$: $\vec{w}^*$ is a strict local maximizer

$\boldsymbol{H}$ is indefinite: $\vec{w}^*$ is a saddle point

$\boldsymbol{H} \succeq 0$ or $\boldsymbol{H} \preceq 0$: $\vec{w}^*$ can be any of the above three kinds of stationary point, or a local minimizer, or a local maximizer

6.3 Examples of Functions with a 2D Vector Argument

What is the "shape" of a function that has a Hessian matrix with mixed-sign or zero eigenvalues?

For a function with a vector argument that has 2 entries – in elementary calculus, for a function of two variables – the characterizations of stationary points by the eigenvalues of the Hessian matrix is modestly complicated. The Hessian matrix is a 2×2 matrix so there are three characterizations that we can explore:

- Indefinite eigenvalues, one $\lambda_i > 0$ and one $\lambda_j < 0$
- Semidefinite eigenvalues, one $\lambda_i > 0$ and one $\lambda_j = 0$
- Null eigenvalues, both $\lambda_i = 0$ and $\lambda_j = 0$

In the definite characterization, the Hessian matrix has two positive eigenvalues. Geometrically, this implies that there the function $f(\vec{w})$ is convex in any direction. A traditional example is the "bowl" function

$$f(\vec{w}) = w_1^2 + w_2^2 \tag{6.4}$$

A surface plot of the function defined in Equation 6.4 is shown in Figure 6.1.

In the indefinite characterization, the Hessian matrix has one positive eigenvalue and one negative eigenvalue. Geometrically, this implies that there is a direction in which the function $f(t)$ is convex and a direction in which the function $f(t)$ is concave. A traditional example is the "saddle" function

$$f(\vec{w}) = w_1^2 - w_2^2 \tag{6.5}$$

A surface plot of the function defined in Equation 6.5 is shown in Figure 6.2.

In the semidefinite characterization, the Hessian matrix has one positive eigenvalue and one zero eigenvalue. Geometrically, this implies that there is a direction in which the function $f(t)$ is

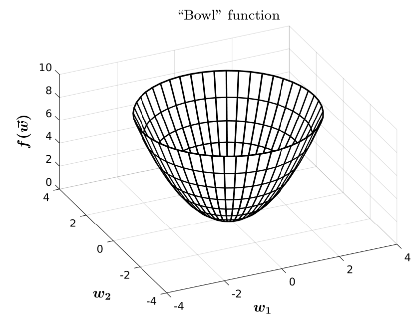


Figure 6.1: The "bowl" function has, at the stationary point $\vec{w}^* = \vec{0}$, a Hessian matrix with two positive eigenvalues.

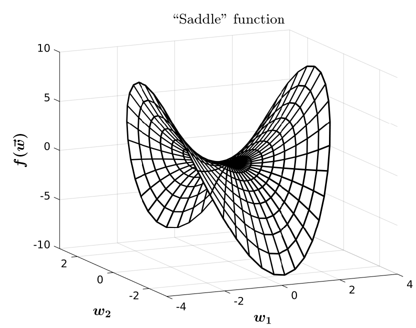


Figure 6.2: The "saddle" function has, at the stationary point $\vec{w}^* = \vec{0}$, a Hessian matrix with one positive eigenvalue and one negative eigenvalue. The positive eigenvalue corresponds to the direction in which $f(\vec{w})$ increases positively without limit. The negative eigenvalue corresponds to the direction in which $f(\vec{w})$ decreases negatively without limit.

convex and a direction in which the function $f(t)$ is flat – locally, neither increasing nor decreasing. One example looks like a piece of a tortilla that is gently curved along an axis, such as

$$f(\vec{w}) = (w_1 + w_2)^2 \tag{6.6}$$

A surface plot of the function defined in Equation 6.6 is shown in Figure 6.3.

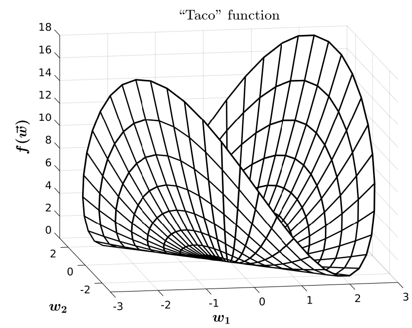


Figure 6.3: The "taco" function has, at the stationary point $\vec{w}^* = \vec{0}$, a Hessian matrix with one positive eigenvalue and one zero eigenvalue. The positive eigenvalue corresponds to the direction in which $f(\vec{w})$ increases positively without limit. The zero eigenvalue corresponds to the direction in which $f(\vec{w})$ is constant.

In the null-matrix characterization, the Hessian matrix has both eigenvalues equal to zero. Geometrically, this implies that the function $f(x)$ is locally flat in all directions. One example is the "monkey saddle" function, which can be specified as

$$f(\vec{w}) = w_1(w_1^2 - 3w_2^2) \tag{6.7}$$

A surface plot of the function defined in Equation 6.6 is shown in Figure 6.4.

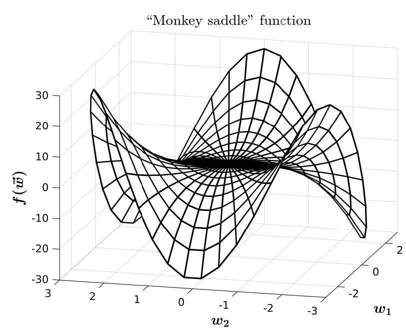


Figure 6.4: The "monkey saddle" function has, at the stationary point $\vec{w}^* = \vec{0}$, a Hessian matrix with both eigenvalues equal to zero. Locally, the function is flat at the stationary point. There are 3 directions is which $f(\vec{w})$ increases positively without limit and 3 directions is which $f(\vec{w})$ decreases negatively without limit.

A useful geometrical construction for a 2D stationary point $\vec{w}^*$ is to create an infinite line in the 2D plane that (a) contains the stationary point and (b) has a given direction vector $\vec{v}$. We can write such a line, using vector notation, as

$$\vec{l}(t) = \vec{w}^* + t\vec{v}$$

We can use a point on such a line as the argument to an objective function. This is the composition of functions

$$f(\vec{l}(t))$$

The composition can be used to visualize a cross-section of the objective function in a vertical plane that is perpendicular to the horizontal coordinate plane and that contains the line $l(\mu)$. This cross-sectioning is particularly powerful when we choose a line direction that is an eigenvector of the Hessian matrix of the objective function at the stationary point $\vec{w}^*$ of interest.

The function described by Equation 6.5 has an indefinite Hessian matrix. The cross-section through the stationary point that is in the direction of the negative eigenvector is a concave parabola. The cross-section through the stationary point that is in the direction of the positive eigenvector is a convex parabola. These curves are shown in Figure 6.5.

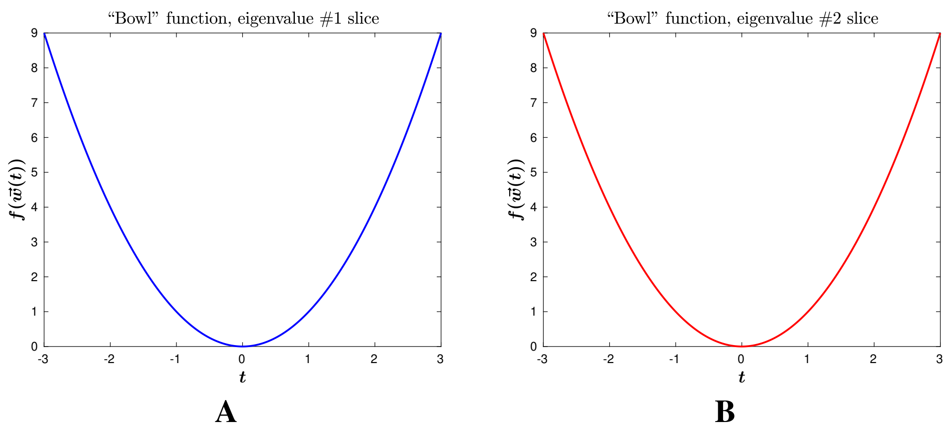
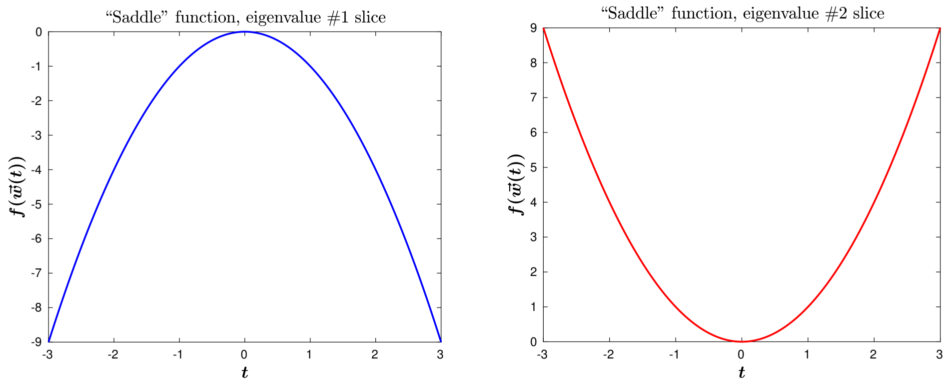


Figure 6.5: Cross-sections of the "saddle" function, through the stationary point and in the directions of the eigenvectors of the Hessian matrix at the stationary point. (A) The direction for the negative eigenvalue is a concave parabola. (B) The direction for the positive eigenvalue is a convex parabola.

The function described by Equation 6.6 has a positive semidefinite Hessian matrix. The cross-section through the stationary point that is in the direction of the null eigenvector is a "flat" line. The cross-section through the stationary point that is in the direction of the positive eigenvector is a convex parabola. These curves are shown in Figure 6.6.

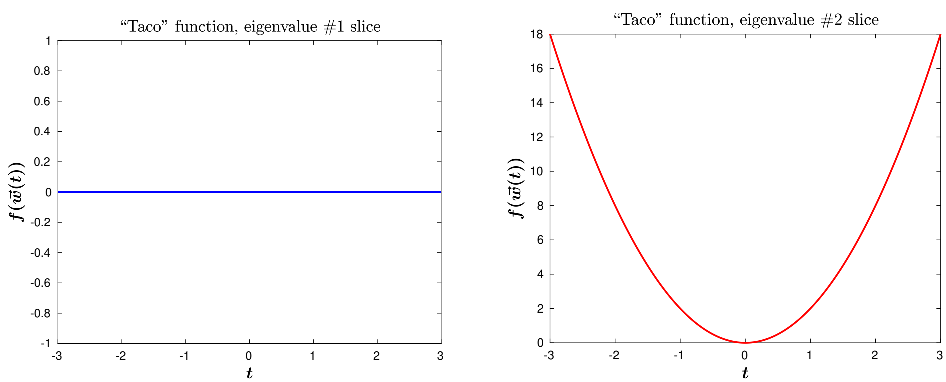
The function described by Equation 6.7 has a Hessian matrix with zero in each entry; both eigenvalues are zero and, by convention, the eigenvectors are elementary vectors that form an orthogonal basis for the 2D plane. The cross-section through the stationary point that is in the direction of the first eigenvector is a cubic curve with a "flat" inflection at the stationary point.



A **B**

Figure 6.6: Cross-sections of the "taco" function, through the stationary point and in the directions of the eigenvectors of the Hessian matrix at the stationary point. (A) The direction for the null eigenvalue is a constant function, drawn as a horizontal line. (B) The direction for the positive eigenvalue is a convex parabola.

The cross-section through the stationary point that is in the direction of the second eigenvector is a "flat" line. These curves are shown in Figure 6.7.



A **B**

Figure 6.7: Cross-sections of the "monkey saddle" function, through the stationary point and in the directions of the eigenvectors of the Hessian matrix at the stationary point. (A) The direction for the first eigenvalue is a cubic that has an inflection at the stationary point. (B) The direction for the second eigenvalue is a constant function, drawn as a horizontal line.

If the Hessian matrix is semidefinite, then the function $f(\vec{w})$ is locally flat at the stationary point $\vec{w}^*$ in an important sense: $\vec{w}^*$ might be a local minimizer but it is not a *strict* local minimizer, because there is a direction in which $f(\vec{w})$ does not change its "height" or value.

If the Hessian matrix is indefinite, then the function $f(\vec{w})$ has a curious local curvature at the saddle point $\vec{w}^*$: in at least one direction the function $f(\vec{w})$ is convex or "up', and in at least one direction the function $f(\vec{w})$ is concave or "down". This provides us with a sufficient condition for a stationary point to be a saddle point.

Extra Notes

6.4 Extra Notes on the Hessian Matrix

To understand the second-order characterization of a stationary point, we can use a simple theorem and its variant.

Theorem: strict minimizer and a symmetric matrix

For any $K \in (\mathbb{R}^n) \times (\mathbb{R}^n) : K = K^T$, for the function $g(\vec{v}) \stackrel{\text{def}}{=} \vec{v}^T K \vec{v}$,

$$(K \succ 0) \rightarrow (\vec{0} \text{ is a strict local minimizer of} g) \tag{6.8}$$

Proof: By expansion, $g(\vec{0}) = 0$. For any $\vec{v} \neq \vec{0}$, $(K \succ 0) \rightarrow (g(\vec{v}) > 0)$. Therefore, $\vec{0}$ is a strict minimizer of g.

Observation: A similar argument shows that if $K \prec 0$, then $\vec{0}$ is a global maximizer of g.

Together, these results justify the statements regarding the definite nature of the Hessian matrix and how a stationary point is a local optimizer of a function. If we take an infinitesimally small step in any direction from such a stationary point of a function, then a second-order approximation to the function will "look" like the quadratic model g.

End of Extra Notes

References

[1] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

[2] Antoniou A, Lu WS: Practical Optimization: Algorithms and Engineering Applications. Springer Science & Business Media, 2007